3Т

УДК 519.863.001.63

В.А. Тазетдінов
Черкаський державний технологічний університет, м. Черкаси, Україна
valeriy_tazetdinov@rambler.ru

Ідентифікація функції оцінки об'єкта нерухомості нейронною мережею

У статті викладено аспекти визначення ціни об'єкта нерухомості з використанням модифікованого алгоритму стохастичної релаксації. Запропонований метод ідентифікації функції оцінки об'єкта нерухомості впроваджено в інформаційно-аналітичній системі «REMA». Виконано верифікацію отриманих результатів з використанням штучних нейронних мереж.

Вступ

Інтерес до нейромережних технологій, який виявляють фахівці з різних сфер діяльності, пояснюється, насамперед, дуже широким діапазоном розв'язуваних з їхньою допомогою задач, а також перевагами перед іншими методами.

Аналіз робіт, пов'язаних із використанням нейронних мереж для аналізу ринку нерухомості, а саме: розв'язання задач класифікації, ідентифікації, прогнозування та розпізнавання образів, показує, що нейромережний підхід має переваги перед традиційними математичними методами в трьох випадках. По-перше, коли задача, що розглядається, через конкретні особливості предметного середовища не може бути адекватно формалізована, оскільки містить елементи невизначеності, які не формалізуються з використанням традиційних математичних понять. По-друге, якщо задачу можна формалізувати, але на даний час апарат для її розв'язання відсутній. По-третє, коли для формалізованої задачі існує відповідний математичний апарат, але реалізація обчислень з його допомогою на базі наявних обчислювальних систем не задовольняє вимогам одержання розв'язку за критеріями часу та іншими. У такій ситуації доводиться робити спрощення алгоритмів, що знижує якість рішень, або застосовувати відповідний нейромережний підхід за умови, що він забезпечить потрібну якість виконання задачі.

Постановка задачі дослідження

Необхідність врахування значної кількості екзогенних факторів вимагає достатньої кількості навчальних образів. Багатофакторність задачі і, як наслідок, рельєфність функції помилки вказують на наявність локальних мінімумів і високої ймовірності влучення в них. Існує також небезпека паралічу мережі [1].

Уникнути проблем, пов'язаних із навчанням нейронної мережі НМ, дозволяє процедура навчання Больцмана [1], [2]. Її головна ідея полягає у використанні принципу віджигу металу. Якщо метал нагріти до температури, яка перевищує його точку плавлення, то атоми знаходяться у стані невпорядкованого руху. При охолодженні вони прагнуть до стану, який відповідає мінімуму енергії. Ймовірність того, що система знаходиться у стані з енергією e, визначається розподілом Больцмана

$P(e)=\exp\left(-\frac{e}{kT}\right)$, де k – постійна Больцмана, T – температура. Головний висновок із цієї формули полягає в тому, що високоенергетичні та низькоенергетичні стани майже рівноймовірні. При високих температурах ймовірність прагне до одиниці незалежно від енергетичного стану. При наближенні температури до нуля ймовірність високоенергетичного стану є близькою до нуля. Ці принципи і покладені в основу стохастичної релаксації НМ.

Модифікований алгоритм методу стохастичної релаксації (МАСР) [3] є таким:

Крок 1. Задати початкове, заздалегідь велике значення температури T і точність результату ε.

Крок 2. Згенерувати рівномірно розподілені на інтервалі (0, 1) матриці значень вагових коефіцієнтів $W=(w_{pq})_{p,q=1}^{n\ u}, V=(v_p)_{p=1}^{u}$. Нехай $i=1,\ p=1,\ q=1$.

Крок 3. Подати на вхід мережі i-й навчальний образ і всі контрольні образи та обчислити значення функції енергії E_1 і E_k, відповідно.

Крок 4. Змінити значення вагового коефіцієнта $w_{pq}\in W$ на рівномірно розподілену в (0, 1) величину Δw_{pq}.

Крок 5. Подати на вхід мережі i-й навчальний і всі контрольні образи та обчислити значення функції енергії $E_1^{'}$ і $E_k^{'}$.

5.1. Якщо $E_1^{'}\le E_1$ і $E_k^{'}\le E_k$, то зміну вагового коефіцієнта зберегти.

5.2. Якщо $E_1^{'}> E_1$ і $E_k^{'}> E_k$, то зміну скасувати.

5.3. Якщо $E_1^{'}> E_1$ і $E_k^{'}\le E_k$, то генеруємо рівномірно розподілене на (0,1) випадкове число ζ та знаходимо $F_2^{-1}(\Delta w_{pq})$. При виконанні нерівності $\zeta > F_2^{-1}(\Delta w_{pq})$ зміну вагового коефіцієнта зберігаємо, у протилежному випадку значення вагового коефіцієнта залишаємо незмінним.

5.4. Якщо $E_1^{'}\le E_1$ і $E_k^{'}> E_k$, то процедура зміни значення вагового коефіцієнта аналогічна п. 5.3, лише використовується функція $F_1^{-1}(\Delta w_{pq})$.

Крок 6. Якщо подані всі навчальні образи, то розрахувати значення цільової функції E на всіх навчальних і контрольних образах. Якщо $E<\varepsilon$, то перейти на крок 9.

Крок 7. Якщо здійснено перебір усіх вагових коефіцієнтів із матриць W і V, зменшити значення температури, $i=i+1$. Перейти на крок 3.

Крок 8. В якості w_{pq} взяти наступний ваговий коефіцієнт із W або V і перейти на крок 4.

Крок 9. Кінець.

Запропонований алгоритм майже унеможливлює виникнення традиційних для НМ із градієнтними методами навчання проблем із влученням у локальні мінімуми функції енергії і «паралічем» мережі. Відмінністю запропонованого методу від класичного навчання Больцмана є використання принципу регуляризації [4], відповідно до якого всі дані (крім перевірочної послідовності) певним чином поділяються на дві послідовності: навчальну і контрольну.

Отже, необхідно впровадити розроблений алгоритм в інформаційно-аналітичну систему і дослідити ефективність його використання для визначення функції оцінки об'єкта нерухомості (ОН).

Ідентифікація функції оцінки об'єкта нерухомості нейронною мережею із модифікованим алгоритмом стохастичної релаксації

Визначення функції оцінки ОН будемо виконувати за допомогою інформаційно-аналітичної системи ІАС REMA (Real Estate Management Analysis) [5]. Для розв'язання задачі ідентифікації функції ціни ОН в ІАС REMA передбачено два режими роботи: з ручним та автоматичним вибором параметрів. Базовою моделлю є НМ із прямозв'язним функціонуванням. Структура НМ визначається кількістю входів, виходів, прихованих шарів, нейронів у прихованих шарах. Якщо кількість входів та виходів однозначно визначається формалізованою постановкою задачі, то вибір параметрів прихованих шарів залишається досі не вирішеною проблемою. Існують лише певні оцінки, які звужують коло пошуку, але однозначної відповіді не дають.

При автоматичному виборі параметрів використані такі міркування. Оскільки у відомій теоремі Колмогорова [6] встановлено, що функція k-змінних може бути представлена як суперпозиція $2k+1$ одновимірних функцій, то недоцільно вибирати кількість нейронів прихованого шару більшою ніж подвоєне число вхідних факторів, тобто $1 \le p \le 2n_x + 1$. З іншого боку, показано [7], що, припускаючи границі помилки [0, 1/8], кількість навчальних образів повинна бути приблизно рівною кількості вагових коефіцієнтів НМ, помножена на обернену величину помилки, тобто $m \ge \frac{N}{\varepsilon}$. Тоді загальна кількість вагових коефіцієнтів $N = p \cdot n_{x+y}$, де n_{x+y} – сумарна кількість входів і виходів НМ і $n_{x+y} \le N \le (2n_{x+y}+1)n_{x+y}$. Результуючий вираз для обчислення кількості нейронів прихованого шару є таким:

$$p \in [1, \frac{1}{n_{x+y}} \min\{m\varepsilon, (2n_x+1)n_{x+y}]. \quad (1)$$

Для 500 навчальних образів, точності результату 0,1, 4-х входів та 1-му виходу максимальна кількість нейронів прихованого шару становить 9. Всі вищенаведені висновки, припущення та результати одержані без врахування того факту, що апроксимуючі функції, які використовуються як активаційні у нейронах НМ є різними залежно від задачі та типу і нормування початкових даних.

Визначення оптимальної кількості нейронів прихованого шару здійснимо експериментально за допомогою ІАС REMA. Використаємо процедуру «ручного» вибору початкових значень параметрів НМ та алгоритму її навчання. Нехай:

– кількість нейронів прихованого шару – від 3 до 100;
– вхідних нейронів – 4;
– вихідних нейронів – 1;
– значення середньої абсолютної помилки для навчальної нормованої послідовності – від 2 до 0,15;
– точок навчальної послідовності – 39.

Результати роботи ІАС наведені на рис. 1 та в табл. 1 (e – середня абсолютна помилка, n – кількість нейронів прихованого шару). Їх аналіз показує, що мінімальним є середнє абсолютне відхилення на контрольній послідовності, одержане при мінімальному значенні середнього абсолютного відхилення на точках навчаль-

ної послідовності, що є очікуваним. Разом із тим встановлено, що оптимальною є кількість нейронів прихованого шару від 6 до 10, що суперечить лише (1).

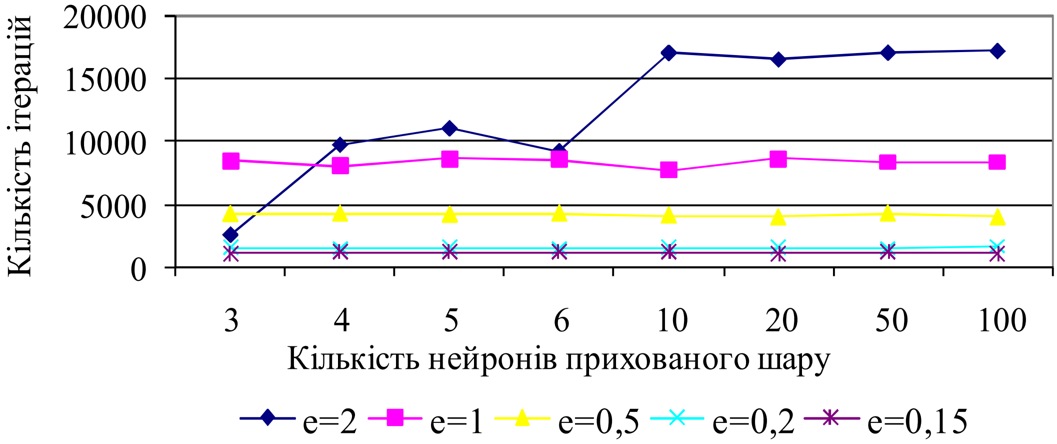


Рисунок 1 – Результати експерименту

Таким чином, одержані результати свідчать про необхідність перенавчання НМ у випадку збільшення кількості точок навчальної послідовності на декілька відсотків (4 – 8 %). Виконувати цю процедуру раціонально щомісячно.

Таблиця 1 – Кількість ітерацій алгоритму МАСР

e\n	3	4	5	6	10	20	50	100
2	2507	9666	10981	9162	16998	16471	17002	17144
1	8382	8022	8584	8501	7652	8551	8262	8300
0,5	4253	4294	4204	4277	4120	3977	4273	3985
0,2	1522	1500	1538	1502	1537	1527	1485	1594
0,15	1091	1174	1152	1189	1159	1057	1226	1108

Для того щоб переконатись у ефективності роботи МАСР (з використанням ІАС REMA), виконаємо експериментальну перевірку та порівняння результатів роботи НМ із алгоритмами Левенберга-Маркарта (АЛМ), алгоритмом оберненого поширення похибки із оптимізацією вагових коефіцієнтів методом Флетчера-Пауела (АФП), алгоритмом спряжених градієнтів (АСГ) (з використанням пакету Matlab). Оскільки кожний із цих алгоритмів реалізує прямозв'язна НМ, то параметри структури задані однаково. Початковими даними є вибірка ОН, яка використовувалась у попередній процедурі. В якості критеріїв вибрано:

– час навчання НМ (при однаковому значенні середнього абсолютного відхилення на точках навчальної послідовності);

– величина середнього абсолютного відхилення точних значень результуючого показника від розрахованих НМ на точках контрольної послідовності (при фіксованому значенні такої ж величини на точках навчальної послідовності);

– величина середнього абсолютного відхилення точних значень результуючого показника від розрахованих НМ на точках контрольної послідовності (при фіксованому значенні часу навчання НМ).

Результати експерименту наведені в табл. 2. Для їх верифікації розглядались три вибірки із генеральної сукупності. У першій вибірці відношення навчальних образів до контрольних становило 80:20, у другій – 70:30, у третій – 60:40. Для розрахунку значення другого критерію К2 значення середнього абсолютного відхилення

на точках (нормованих) навчальної послідовності встановлено рівним 0,3. Знаходження значення К3 відбувалось при значенні середнього абсолютного відхилення на точках навчальної послідовності 0,0001, що апріорі вимагало додаткових перетворень для досягнення такої точності на фіксованому часі навчання, рівному одній хвилині.

Таблиця 2 – Дані результатів експерименту

Номер вибірки	НМ	Критерії		
		К1	К2	К3
1	МАСР	62	68	128
	АЛМ	31	251	278
	АФП	27	208	245
	АСГ	38	284	259
2	МАСР	61	82	135
	АЛМ	27	305	288
	АФП	26	296	270
	АСГ	39	310	272
3	МАСР	56	128	138
	АЛМ	26	351	306
	АФП	20	329	305
	АСГ	31	362	338

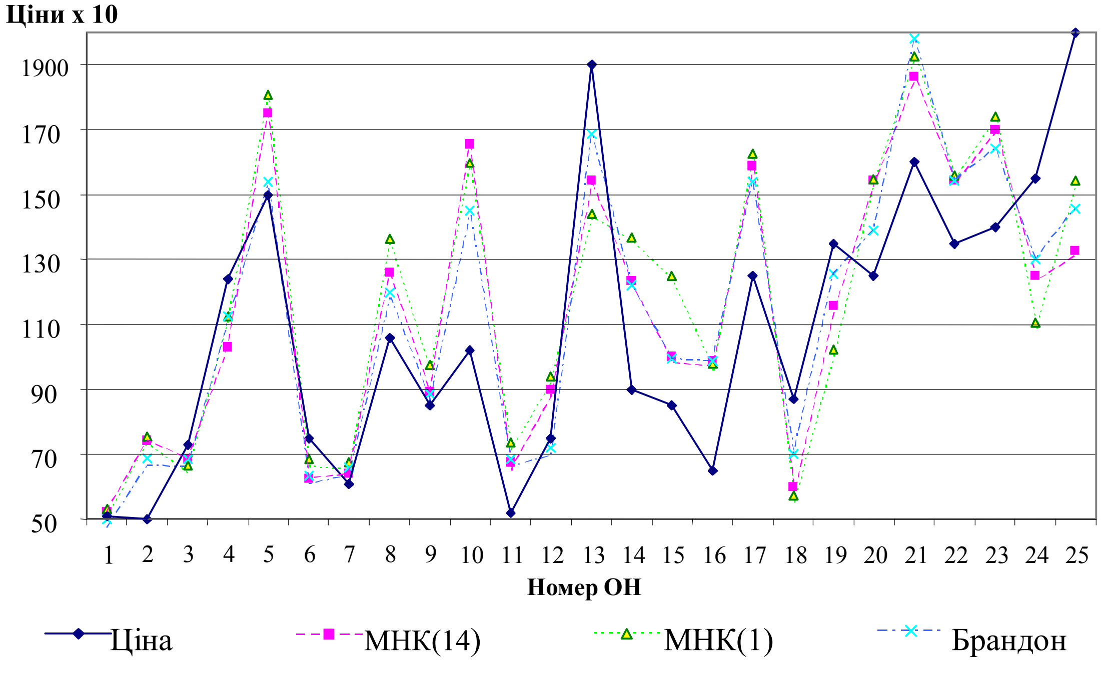


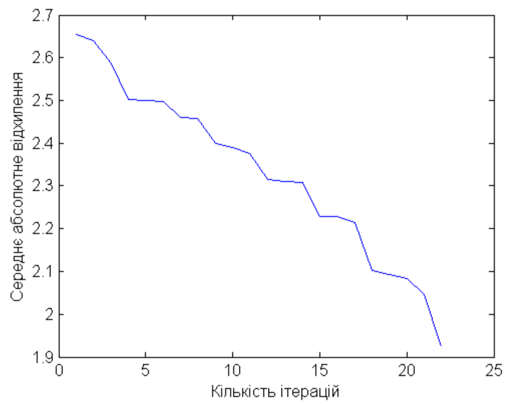
Рисунок 2 – Реальна ціна і ціна, розрахована з використанням нейромережної ідентифікації

Під час навчання НМ, в яких використовувалось обчислення градієнта, досить часто спостерігалось явище паралічу мережі, що пояснюється великою кількістю однакових значень факторів. Наслідком цього є нульове значення градієнта та відсутність динаміки навчання НМ. Так, при 50 запусках одного із представників back

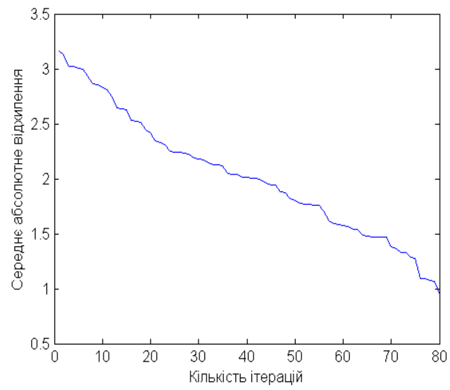
propagation – алгоритму Левенберга-Маркуарда 28 раз процес навчання переривався штучно через параліч мережі, 16 раз точність була низькою через попадання в локальні мінімуми і лише 6 раз точність порівнянна із точністю результатів нашого методу.

Аналіз результатів експериментів показує, що алгоритм МАСР в середньому удвічі повільніше працює за інші алгоритми з градієнтними методами навчання. В той же час його результати є значно точнішими (на 80 – 180 %) як за результати роботи НМ з іншими алгоритмами, так і за результати (800 – 1000 %), що одержані внаслідок використання класичних інтегро-диференціальних методів. При зроблених попередніх припущеннях середня похибка для МАСР становить 0,05 – 0,98 %, для АЛМ – 2,88 – 3,3 %, для АФП – 2,22 – 2,79 %, для АСГ – 3 – 3,39 %, що переконує в ефективності першого алгоритму (рис. 2).

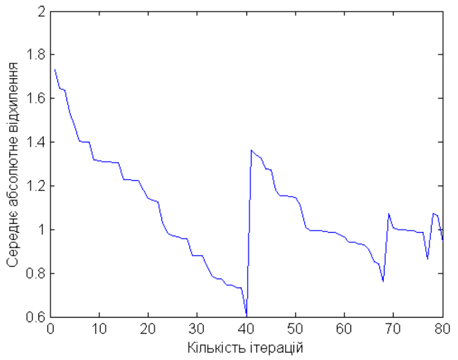
Було проведене також додаткове дослідження графіка залежності середнього квадратичного відхилення від кількості ітерацій [8] (рис. 3). Встановлено, що досить велике значення середньої абсолютної похибки (Error = 2, Error = 1,5, Error = 1) на точках контрольної послідовності зумовлює досить гладке зменшення середньої абсолютної похибки на точках контрольної послідовності. При зменшенні значення похибки на точках навчальної послідовності в середньому після 30 – 50 ітерацій середня абсолютна похибка на точках контрольної послідовності змінюється стрибкоподібно.



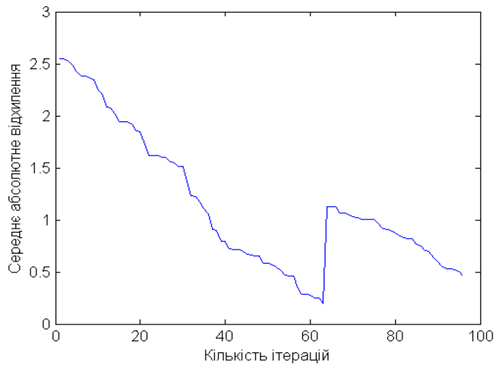
а) Error = 2



б) Error = 1



в) Error = 0,7



г) Error = 0,2

Рисунок 3 – Графіки залежності середньої абсолютної похибки на контрольній послідовності від кількості ітерацій для різних значень середньої абсолютної похибки на навчальній послідовності

Важливо помітити, що є сенс запам'ятовувати значення вагових коефіцієнтів, які відповідають мінімальним значенням цільової функції. Оскільки важливою складовою МАСР є випадкова зміна значень матриці вагових коефіцієнтів, то неможливо стовідсотково гарантувати збіжність такого методу. І навіть велика кількість ітерацій не завжди приводить до бажаного результату. Запам'ятовування оптимальних значень вагових коефіцієнтів дозволить відновити функціонування НМ із точнішою ідентифікацією шуканої залежності.

Висновки

Як відомо, суб'єктивні переваги людини, особливо при виборі і купівлі нерухомості, мають значну кількість «локальних екстремумів». Так, якщо покупця влаштовує загальна площа квартири, то він може не звернути уваги на її недоліки, хоча міг би вибрати і більш збалансований варіант. Проблеми такого роду і допоможе вирішити НМ із модифікованим алгоритмом стохастичної релаксації. Знаходження оптимального варіанта і визначення реальної ціни – задачі, які така НМ допомагає розв'язувати якнайкраще.

Запропонований метод ідентифікації функції оцінки ОН впроваджено в інформаційно-аналітичній системі «REMA», яка дозволяє здійснювати аналіз тенденцій, що складаються на ринку, прогнозувати динаміку зміни ціни ОН. Як показують отримані результати, використання розробленого методу дозволить розв'язувати задачу визначення функції оцінки об'єктів нерухомості з більш високою ефективністю.

Література

1. Уоссермен Ф. Нейрокомпьютерная техника: теория и практика. – М.: Мир, 1992. – 190 с.
2. Geman S., Geman D. Stochastic relaxation, Gibbs distribution and Baysian restoration of images // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1984. – Vol. 6. – P. 721-741.
3. Снитюк В.Е., Тазетдинов В.А. Применение метода стохастической релаксации для прогнозирования рынка недвижимости // Сборник докладов Междунар. научн. конф. «Нейросетевые технологии и их применение»: (Краматорск). – 2003. – С. 226-236.
4. Ивахненко А.Г. Долгосрочное прогнозирование и управление сложными системами. – К.: Техника, 1975. – 312 с.
5. Тазетдінов В.А. Структура програмно-алгоритмічного забезпечення процесів прийняття рішень на ринку нерухомості // Тези доп. учасників VI Міжнар. наук.-практ. конф. студентів, аспірантів та молодих вчених «Системний аналіз та інформаційні технології». – К.: НТУУ «КПІ», 2004. – С. 134-136.
6. Колмогоров А.Н. О представлении непрерывных функций нескольких переменных в виде суперпозиции непрерывных функций одного переменного // Докл. АН СССР. – 1957. – Т. 114, № 5. – С. 679-681.
7. Горбань А.Н. Обучение нейронных сетей. – М.: Изд. СССР-США СП «ParaGraph», 1990. – 160 с.
8. Гультяев А.К. Визуальное моделирование в среде MATLAB: Учеб. курс. – СПб.: Питер, 2000. – 432 с.

В.А. Тазетдинов
Идентификация функции оценки объекта недвижимости нейронной сетью
В статье изложены аспекты определения цены объекта недвижимости с использованием модифицированного алгоритма стохастической релаксации. Предложенный метод идентификации оценки объекта недвижимости внедрен в информационно-аналитическую систему «REMA». Выполнена верификация полученных результатов с использованием искусственных нейронных сетей.

V.A. Tazetdinov
Identification of Cost Function of Real Estate Object with Neural Net
The aspects of determination of cost of real estate objects with usage of tutoring stochastic relaxation modified algorithm have been developed. This method of identification of cost of real estate object has been introduced in information analytical system «REMA». The verification of obtained results with usage of neural artificial nets has been accomplished.

Стаття надійшла до редакції 27.08.2008.